Space-Time Models with Time Dependent Covariates[1]

Modelli spazio-tempo con covariate tempo dipendenti

Emanuela Dreassi
Dip. di Statistica "G. Parenti"
Università di Firenze
dreassi@ds.unifi.it

Annibale Biggeri
Dip. di Statistica "G. Parenti"
Università di Firenze
abiggeri@ds.unifi.it

Dolores Catelan
Dip. di Statistica "G. Parenti"
Università di Firenze
catelan@ds.unifi.it

Corrado Lagazio
Dip. di Scienze Statistiche
Università di Udine
lagazio@dss.uniud.it

Riassunto: Viene analizzato il legame tra istruzione e mortalità per tumore al polmone nella popolazione maschile della Toscana, a livello comunale, nel periodo 1971-99, diviso in periodi quinquennali. In particolare, fissate quattro rilevanti età all'esposizione, si intende stimare quale fra esse è maggiormnete correlata con la mortalità nei diversi periodi facendo uso di un modello spazio-temporale Bayesiano con effetti di coorte e covariate tempo dipendenti.

Keywords: Space-time hierarchical Bayesian models, Time-dependent covariates.

1. Introduction

The aim of ecological studies is to describe the relationship between geographical variation of disease risk and concomitant variation in the level of exposure to a particular factor: for example, an environmental agent or a life-style related characteristic. In this kind of research it is often necessary to account for the relevant aetiological period between exposure and the occurrence of the disease. With regard to mortality for lung cancer the biology of the process of carcinogenesis suggests that more than 10-15 years should run between exposure and mortality. A space-time model should then be specified in order to allow for this latency period; see for example Dreassi (2003) and Dreassi *et al.* (2005). An additional problem arise when studying lung cancer mortality at ecological level: we rarely have information at aggregate level on smoking habits, one of the most important determinants of the aetiology of the disease. We must rely on some proxies of this factors. In the literature two are the most common variables used to summarize socioeconomic factors: deprivation index and education. The first (the variable used on the previous cited paper) reflects the prevalence of subject characteristics such as unemployment, low education, living in a small dwelling, overcrowding, not having a car; see Townsend *et al.* (1988). The second one can be used both as proxy of income or to reflect life style behavior. In particular, recent studies have shown that there is a relation between

[1]The research was partially supported by COFIN-MIUR 2004 and SLTo-Tuscany Region Project. We are grateful to Dott.ssa Mariangela Vigotti (University of Pisa), Dott.ssa Elisabetta Chellini (Center of Study on Prevention of Cancer, CSPO, Florence and Regional Mortality Register) and Dott.ssa Paola Baldi (Regione Toscana) for having kindly made available the data used in the present work.

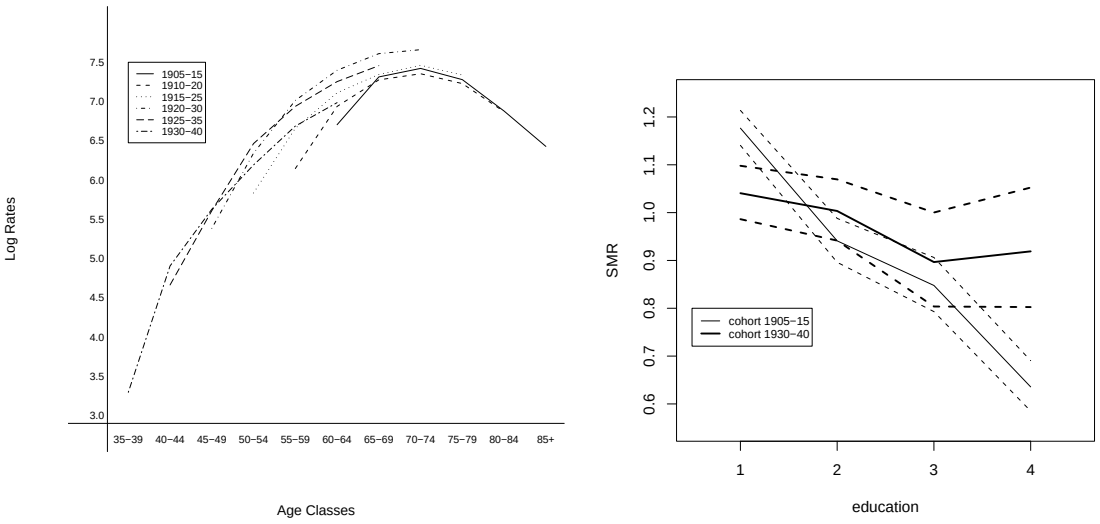
level of education and smoking consumption (see, for example, Cavelaars *et al.*, 2000). Following this field of research, in this work we use education as covariate that could explain the space-time pattern of the disease and time dimension is the birth-cohort (see Lagazio *et al.*, 2001 and Lagazio *et al.*, 2003). The main goal of the model is to identify the relevant age at exposure in the relation between a proxy of socioeconomic factors (e.g. level of education) and lung cancer mortality from the birth cohort 1910-15 to cohort 1930-40 in males in Tuscany (Italy). A hierarchical Bayesian model with time-dependent covariates, latency periods, time and space random terms, and time misaligned exposure-disease data is proposed. Results confirm the presence of an association between mortality for lung cancer and socioeconomic factors with a relevant age at exposure between 20 and 30.

In Section 2 we describe the data. In Section 3 we present standard models that take into account only the space dimension, then we introduce different hierarchical Bayesian space-time models with time-dependent covariates. Results obtained with the different models are shown in Section 4. In Section 5 we discuss the results, the proposed space-time models, their limits, and their possible extensions.

2. Data

Lung cancer death certificates are considered for males resident in the 287 municipalities of the Tuscany Region (Italy) from 1971 to 1999. For the 29 years analyzed, amounting to a total of 49,684,302 person-years, the number of recorded death certificates was 47,343. Data were made available by the Tuscany Regional Government under the research project *Tuscany Atlas of Mortality 1971-1994* (see Vigotti *et al.*, 2001) and by the Regional Mortality Register for the period 1995-99. Deaths and corresponding populations for each municipality were cross classified by 18 age classes (0-4, 5-9, 10-14, . . ., 85 and more) and 6 calendar periods (1971-74, 1975-79, 1980-84, . . ., 1995-99). The expected number of cases in each municipality have been evaluated using the age-specific reference rates calculated using an age-cohort model (see Clayton and Schifflers, 1987). We have used this model to clean the age effect from a cohort component present in the mortality data. In this way we can consider the cohort effect as a main effect in the Bayesian specifications. Figure 1a shows the logarithm of the cohort specific mortality rates, Figure1b SMRs with their 95% confidence interval versus quartile of education for the extreme cohorts 1905-15 and 1930-40. The expected cases for each age class and birth cohort in each municipality were then calculated by applying the age-specific reference rates to the age-specific person-years of each area. Observed and expected cases were then aggregated along the diagonals of the Lexis diagram representing the six birth cohorts previously defined, thus collapsing on the age dimension. For the space-time analysis we have considered the six birth cohorts (1905-15, . . . , 1930-40) corresponding to people aged between 35 and 64 at the beginning of the study period (see Lagazio *et al.*, 2001 and Lagazio *et al.*, 2003). These cohorts are those followed up for all the considered calendar periods and with substantial observed number of events. Table 1 describe the relation between the birth cohorts to each age classes and calendar periods.

Data on education have been derived from census data collected by the Italian Statistical Institute (ISTAT) on the years 1921, 1931, 1951, 1961, 1971, 1981 and 1991. Considering that relative earning potentials of educational credentials may differ markedly



(a) (b)

Figure 1: (a) Cohort specific mortality log rates. Tuscany, males, 1971-99. (b) SMRs with their 95% confidence interval versus quartile of education for the extreme cohorts 1905-15 and 1930-40

Table 1: Birth cohorts corresponding to each age class and calendar periods

	Periods					
Age	1971-74*	1975-79	1980-84	1985-89	1990-94	1995-99
0-4	1965-75	1970-80	1975-85	1980-90	1985-95	1990-00
5-9	1960-70	1965-75	1970-80	1975-85	1980-90	1985-95
10-14	1955-65	1960-70	1965-75	1970-80	1975-85	1980-90
15-19	1950-60	1955-65	1960-70	1965-75	1970-80	1975-85
20-24	1945-55	1950-60	1955-65	1960-70	1965-75	1970-80
25-29	1940-50	1945-55	1950-60	1955-65	1960-70	1965-75
30-34	1935-44	1940-50	1945-55	1950-60	1955-65	1960-70
35-39	**1930-40**	1935-45	1940-50	1945-55	1950-60	1955-65
40-44	**1925-35**	**1930-40**	1935-45	1940-50	1945-55	1950-60
45-49	**1920-30**	**1925-35**	**1930-40**	1935-45	1940-50	1945-55
50-54	**1915-25**	**1920-30**	**1925-35**	**1930-40**	1935-45	1940-50
55-59	**1910-20**	**1915-25**	**1920-30**	**1925-35**	**1930-40**	1935-45
60-64	**1905-15**	**1910-20**	**1915-25**	**1920-30**	**1925-35**	**1930-40**
65-69	1900-10	**1905-15**	**1910-20**	**1915-25**	**1920-30**	**1925-35**
70-74	1895-05	1900-10	**1905-15**	**1910-20**	**1915-25**	**1920-30**
75-79	1890-00	1895-05	1900-10	**1905-15**	**1910-20**	**1915-25**
80-84	1885-95	1890-00	1895-05	1900-10	**1905-15**	**1910-20**
85+	1880-90	1885-95	1890-00	1895-05	1900-10	**1905-15**

$\star$ The cohorts related to this period are obtained considering that mortality data are available from 1970.

Table 2: Proportion of "illiterate"(a) population, of person who "can read"(b) and with "primary"(c) school degree in the analyzed censuses

	Census year						
Title	1921	1931	1951	1961	1971	1981	1991
(a)	**27.72**	**18.21**	11.01	7.12	4.19	2.24	1.26
(b)	-	-	17.40	14.24	26.14	18.04	11.28
(a)+(b)	-	-	**28.41**	**21.36**	**30.32**	**20.28**	12.54
(c)	72.28	81.79	62.04	64.88	47.88	43.46	36.11
(a)+(b)+(c)	-	-	90.46	86.25	78.21	63.74	**48.66**

for degrees earned, for example, in 1931 versus 1981 we have constructed an indicator assuming that the meaning of "low level of education" changes over time. In particular for each municipality we have considered: the proportion of illiterate population for 1921 and 1931 censuses; the sum of the proportion of illiterate population and of people who can read but do not have any scholar degree for 1951, 1961, 1971 and 1981 censuses; the sum of the proportion of illiterate population, of people who can read but do not have any scholar degree and of people with a primary school degree for 1991 census. In this way we keep as constant as possible over time the proportion of population with a low level of education (results for different definitions are shown in Table 2).

Since mortality and education are recorded at different time points (5 and 10 year lags respectively) we need to align the two series of data estimating a value of the education score for years 1936, 1941, 1946, 1956, 1966, 1976, 1986, 1996 and for each municipality. Unobserved data have been treated as unknown parameters in the model: we specified a prior distribution on them and, by conditioning on the observed data and using Bayes theorem, we sampled, simultaneously, their values and model parameters in the MCMC algorithm. In this way inferences regarding regression coefficients fully take into account additional uncertainty coming from missing data. The education score in each area has been assumed to follow a Normal distribution with mean that is the sum of two components: a heterogeneity random term and a time-autoregressive random term. As a consequence, the imputation algorithm for missing data considers each area having an education score value influenced by the value taken on the two adjacent periods (one for the extreme periods): this is done by introducing time-autoregressive random effects. Moreover, each value of the education score takes into account for unobserved spatial similarity over the entire study region, via the heterogeneity random term.

3. Space-time models

We describe the space-time pattern of mortality risk for lung cancer for the whole region from the cohort 1905-14 to cohort 1930-39 using hierarchical Bayesian models with structured random effects on space and time dimensions. We adopt the hierarchical Bayesian space-time formulation of Knorr-Held (2000), with or without space-time interaction terms, to estimate the relative risk for each cohort and for each municipality. The number of observed cases in the i-th area ($i = 1, \ldots, 287$) and j-th cohort ($j =$ 1905-15, 1910-20, 1915-25, 1920-30, 1925-35, 1930-40) $O_{i,j}$ are assumed to follow a Poisson

distribution with mean $E_{i,j}\theta_{i,j}$, where $E_{i,j}$ indicates the expected number of cases under indirect standardization and $\theta_{i,j}$ the relative risk. A random effects model is assumed for the logarithm of the relative risk

$$\log(\theta_{i,j}) = u_i + v_i + p_j + \xi_{i,j} \tag{1}$$

The term u_i represents an unstructured spatial variability component whose a priori distribution is assumed to be Normal (μ_u, δ_u), and v_i a structured spatial variability component which is modeled, conditionally on $v_{k\sim i}$ terms ($\sim i$ indicates areas adjacent to i-th ones, $k = 1, \ldots, 287$), as Normal $(\bar{v_i}, \delta_v n_i)$; $\bar{v_i} = \sum_{k\sim i} \frac{v_k}{n_i}$ is the mean of the v values calculated on the areas adjacent to the i-th one and n_i their number (conditional autoregressive model). These terms define the random component of the pure spatial model of Besag *et al.* (1991). The term p_j represents the effect of the j-th cohort whose a priori distribution is assumed to be an autoregressive conditional random term $p_j \sim$ Normal$(\bar{p_j}, \delta_p n_j)$; $\bar{p_j}$ is the mean of the $(j-1)$-th and $(j+1)$-th terms (for extreme cohort the $(j+1)$-th or $(j-1)$-th only) and n_j equal 2 (or 1 for the extreme cohorts). The term $\xi_{i,j}$ represents the space-time interaction, whose prior can be specified in several ways depending on the assumptions about the dependence structure. In our model, we assumed that interaction terms are structured both in space and time. The mean of the conditional distribution of $\xi_{i,j}$ given all the other $\xi_{k,j}$ terms (using again symbol $\sim i$ to define adjacent areas to i-th ones) is the following:

$$\begin{cases} \xi_{i,j+1} + \sum_{k\sim i} \frac{\xi_{k,j}}{n_i} - \sum_{k\sim i} \frac{\xi_{k,j+1}}{n_i} & \text{if } j\text{=1905-15} \\ \xi_{i,j-1} + \sum_{k\sim i} \frac{\xi_{k,j}}{n_i} - \sum_{k\sim i} \frac{\xi_{k,j-1}}{n_i} & \text{if } j\text{=1930-40} \\ \frac{(\xi_{i,j-1} + \xi_{i,j+1})}{2} + \sum_{k\sim i} \frac{\xi_{k,j}}{n_i} - \sum_{k\sim i} \frac{(\xi_{k,j-1} + \xi_{k,j+1})}{2n_i} & \text{otherwise} \end{cases}$$

The precision is $n_i\delta_\xi$ for $t = 1$ or $t = T$ and $2n_i\lambda_\xi$ for $t = 2, \ldots, T-1$. The hyperprior distributions of the precision parameters δ_u, δ_v, δ_p and δ_ξ are assumed to be uninformative Gamma distribution.

Since the covariates considered are area-specific and time-dependent their inclusion in the model is alternative to the specification of an interaction among space and time random effects. A preliminary analysis considered descriptive mortality and education for each time span separately: four age-at-exposure (20, 30, 40 and 50 years old) between each census (observed 1931, 1951, 1961, 1971, 1981, 1991 and imputed 1936, 1941, 1946, 1956, 1976, 1986, 1996) and mortality in birth cohorts 1905-15, 1910-20,. . . , 1930-40 (see Table 3 for a schematic view of the associations between mortality in each cohort and education observed at different censuses and inter-censuses).

Education has been considered introducing a covariate on the pure spatial model of Besag *et al.* (1991). Education for area i observed at census $j + l$ (so education at age l is related with mortality observed in birth cohort j) is labeled as $x_{i,j+l}$. The model becomes

$$\log(\theta_{i,j}) = u_i + v_i + \beta_{j,l}\ x_{i,j+l} \tag{2}$$

where $\beta_{j,l}$ whose a priori distribution is assumed to be a non informative Normal and defines the relationship between mortality in the j-th cohort and education at age $l =$

Table 3: Mortality birth cohorts and corresponding exposure years at different age-at-exposure

	Age-at-exposure			
Cohort	age 20	age 30	age 40	age 50
1905-15	1931	1941	1951	1961
1910-20	1936	1946	1956	1966
1915-25	1941	1951	1961	1971
1920-30	1946	1956	1966	1976
1925-35	1951	1961	1971	1981
1930-40	1956	1966	1976	1986

$20, 30, 40, 50$. We fitted a sequence of models considering more than one time span. The first model considers an unique age-at-exposure and an unique β for each time span

$$\log(\theta_{i,j}) = u_i + v_i + p_j + \beta \ \boldsymbol{x}_{i,j+l} \ \boldsymbol{\lambda} \tag{3}$$

where $\boldsymbol{x}_{i,j+l} = (x_{i,j+20}, x_{i,j+30}, x_{i,j+40}, x_{i,j+50})$ is the vector of the education score for the i-th area observed at the different ages of exposure, $\boldsymbol{\lambda} \sim$ multinomial$(\boldsymbol{\pi}, 1)$ and $\boldsymbol{\pi} = (\pi_{20}, \pi_{30}, \pi_{40}, \pi_{50})' \sim$ Dirichlet$(1, 1, 1, 1)$. The last term represents the vector of probabilities attributable to the different time-lags. The coefficient β is assumed to follow a non informative Normal distribution and modulates the relationship between education and mortality. A second model considers different age at exposure values and an unique β for each time span

$$\log(\theta_{i,j}) = u_i + v_i + p_j + \beta \ \boldsymbol{x}_{i,j+l} \ \boldsymbol{\lambda}_j \tag{4}$$

where, in this case, $\boldsymbol{\lambda}_j \sim$ multinomial$(\boldsymbol{\pi}_j, 1)$ and $\boldsymbol{\pi}_j = (\pi_{j20}, \pi_{j30}, \pi_{j40}, \pi_{j50})' \sim$ Dirichlet $(1, 1, 1, 1)$. A further model considers different age-at-exposure values and β parameters for each time span

$$\log(\theta_{i,j}) = u_i + v_i + p_j + \beta_j \ \boldsymbol{x}_{i,j+l} \ \boldsymbol{\lambda}_j \tag{5}$$

where each coefficient β_j is assume to follow a non informative Normal distribution. The last model considers a single age at exposure but different β parameters for each time span

$$\log(\theta_{i,j}) = u_i + v_i + p_j + \beta_j \ \boldsymbol{x}_{i,j+l} \ \boldsymbol{\lambda} \tag{6}$$

All the models take into account a period of "no-exposure" of 20 years from the birth (the plausible age at which one person can start to smoke); hence mortality on the j-th cohort would result in association with a covariate observed at least at time $j + 20$. For all the models described, the marginal posterior distributions for the parameters of interest are approximated by Monte Carlo Markov Chain methods. We have made use of WinBUGS software (Spiegelhalter *et al.*, 2000) in order to perform the MCMC analysis. As Gamerman (1997) suggests, we have adopted "block updating" algorithm. For each model we have run two independent chains; checks for achieved convergence of the algorithm were performed following Gelman and Rubin (1992). We have used Deviance Information Criterion (DIC) (see Spiegelhalter *et al.*, 2002) to compare between models.

Table 4: $\beta^{\star}$ coefficients from model (2) and their credibility interval (CI 90%)

	Age at exposure			
Cohort	β_{j20}	β_{j30}	β_{j40}	β_{j50}
1905-15	-0.132 (-0.179,-0.086)	-0.191 (-0.240,-0.141)	-0.139 (-0.189,-0.089)	-0.135 (-0.178,-0.091)
1910-20	-0.145 (-0.201,-0.092)	-0.147 (-0.200,-0.093)	-0.148 (-0.202,-0.094)	-0.146 (-0.200,-0.093)
1915-25	-0.107 (-0.161,-0.055)	-0.036 (-0.081, 0.009)	-0.061 (-0.099,-0.023)	-0.007 (-0.043, 0.029)
1920-30	-0.094 (-0.134,-0.054)	-0.094 (-0.135,-0.055)	-0.093 (-0.134,-0.053)	-0.094 (-0.136,-0.055)
1925-35	-0.106 (-0.153,-0.059)	-0.088 (-0.128,-0.048)	-0.039 (-0.078,0.001)	-0.087 (-0.125,-0.048)
1930-40	-0.040 (-0.097,0.014)	-0.040 (-0.096,0.014)	-0.038 (-0.094,0.016)	-0.040 (-0.096,0.015)

$\star$ Model (2) does not include the cohort effect p_j, so the meaning of β coefficients for this model differ from those of model (3)-(6)

4. Results

Lung cancer mortality in Tuscany (males) exhibits a strong increase in the last thirty years (1971-1999). The distribution of the relative risks is highly spatially structured with north-west areas at higher risk. Figure 2 shows the relative risk for lung cancer mortality in six different birth cohorts obtained from the space-time model (1). The primary aim of the following analysis is to assess if differentials on mortality, after removing space and time random effects, could be associated with education. Education scores derived at the municipality level for censuses from 1931 to 1991 highlight areas where earlier industrialization occurred. It clearly appears that education has a strong spatial distribution, with a higher level of education (e.g. lower education score) in the north-west part of the region and on the coast, similar to the pattern of mortality risk; see Figure 3. This motivates the analysis. We first describe the association between mortality and education for each birth cohort and each census using a pure spatial model (model 2) with a single covariate at a given birth cohort. Results are reported in Table 4. Remembering that higher education scores reflect areas with low level of education, the coefficients show an inverse relationship between mortality and low education; marginally the coefficients are higher (in absolute value) in the second birth cohort but it is not possible to clearly identify the relevant age at exposure. Using models (3-6), we have then jointly estimated the degree of association between mortality and education and the weights attributable to the different age at exposure. Tables 5–7,9 report β coefficients and π probabilities associated at each age at exposure for the model (3-6). Constraining β coefficient to be unique, we found contradictory evidence in favor of one single age at exposure (compare model 3 *vs* model 4). On the other side, allowing different β coefficients for each birth cohort, we found an higher evidence in favor of age at exposure lower than 40 (compare model 6 *vs* 5). We performed a robustness analysis for the more complex model (model 5) putting bigger prior emphasis on the first and last age at exposure. The results (Table 8 and Figure 4a) point out that our model seems to be robust for the estimate of β coefficients while little changes are induced for the age at exposure probabilities. Bayesian model selection using DIC (Table 10) shows a poor fit of model (6), while the high complexity of model (5) is well counterbalanced by better goodness-of-fit. Model (1), considering space-time interaction and not the covariate, is never preferable in term of complexity and goodness-of-fit. The inverse relationship between low education score and mortality for lung cancer decreased by birth cohort ($\hat{\beta} = -0.165$ in 1905-15; $\hat{\beta} = -0.035$ in 1925-35) becoming positive in the last birth cohort considered; the age at exposure vary between 20-30 years.

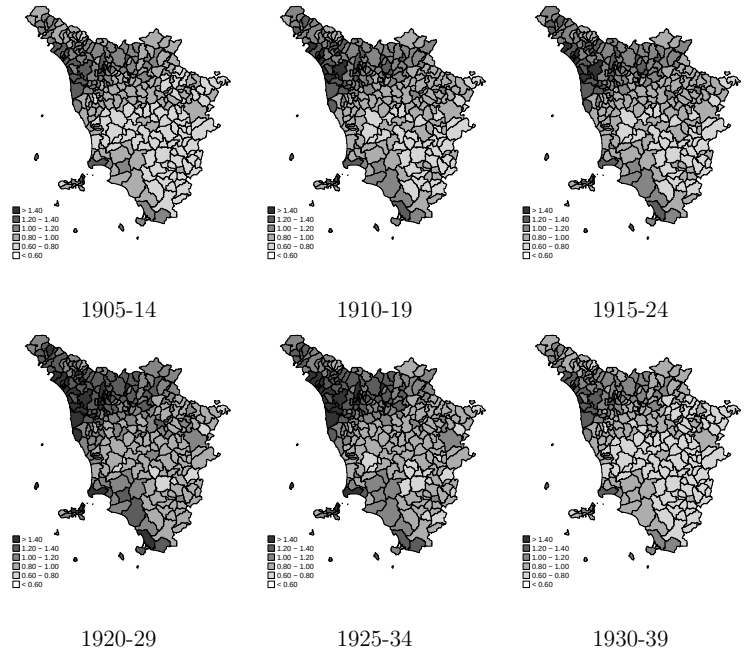


Figure 2: Space-Time distribution of the relative risk for lung cancer, males, Tuscany (Italy). Cohorts 1905-15, . . . , 1930-40

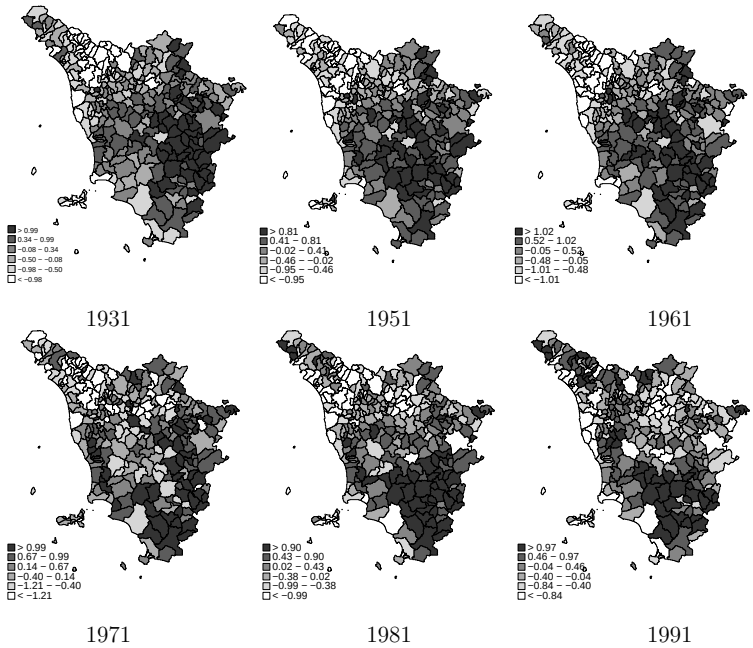


Figure 3: Spatial distribution of the education index in Tuscany (Italy) at censuses considered

Table 5: β coefficient with its credibility interval (CI 90%) and age at exposure probabilities from model (3)

β	π_{20}	π_{30}	π_{40}	π_{50}
-0.098 (-0.119,-0.076)	0.4001	0.2001	0.2002	0.1996

Table 6: β coefficient with its credibility interval (CI 90%) and age at exposure probabilities from model (4)

		cohort	π_{j20}	π_{j30}	π_{j40}	π_{j50}
		1905-15	0.2000	0.4007	0.2007	0.1986
		1910-20	0.3237	0.2748	0.1985	0.2030
β	-0.100 (-0.120,-0.081)	1915-25	0.2071	0.3428	0.2503	0.1997
		1920-30	0.2362	0.2681	0.2319	0.2638
		1925-35	0.3755	0.2237	0.2003	0.2004
		1930-40	0.2514	0.2467	0.2596	0.2513

Table 7: β coefficients with their credibility interval (CI 90%) and age at exposure probabilities from model (5)

cohort	β_j	π_{j20}	π_{j30}	π_{j40}	π_{j50}
1905-15	-0.165 (-0.202,-0.130)	0.2009	0.3989	0.1993	0.2009
1910-20	-0.136 (-0.175,-0.010)	0.2229	0.2005	0.2573	0.3193
1915-25	-0.040 (-0.073,-0.0007)	0.3301	0.2238	0.2403	0.2058
1920-30	-0.054 (-0.092,-0.016)	0.2639	0.2463	0.2302	0.2596
1925-35	-0.035 (-0.072,0.001)	0.2891	0.2649	0.2178	0.2282
1930-40	0.039 (-0.011,0.089)	0.2517	0.2581	0.2456	0.2445

Table 8: β coefficients with their credibility interval (CI 90%) from model (5) when prior distribution uninformative and informative (higher weight on first and last age at exposure) are adopted

cohort	β_j		
	uninformative	first	last
1905-15	-0.165 (-0.210,-0.123)	-0.169 (-0.213,-0.128)	-0.162 (-0.207,-0.118)
1910-20	-0.136 (-0.182,-0.093)	-0.140 (-0.184,-0.098)	-0.132 (-0.179,-0.085)
1915-25	-0.041 (-0.079,0.0007)	-0.048 (-0.084,-0.012)	-0.033 (-0.077,0.016)
1920-30	-0.056 (-0.101,-0.010)	-0.060 (-0.103,-0.017)	-0.051 (-0.097,-0.001)
1925-35	-0.037 (-0.081,0.007)	-0.045 (-0.089,-0.002)	-0.027 (-0.072,0.0155)
1930-40	0.035 (-0.025,0.096)	0.034 (-0.029,0.091)	0.042 (-0.019,0.104)

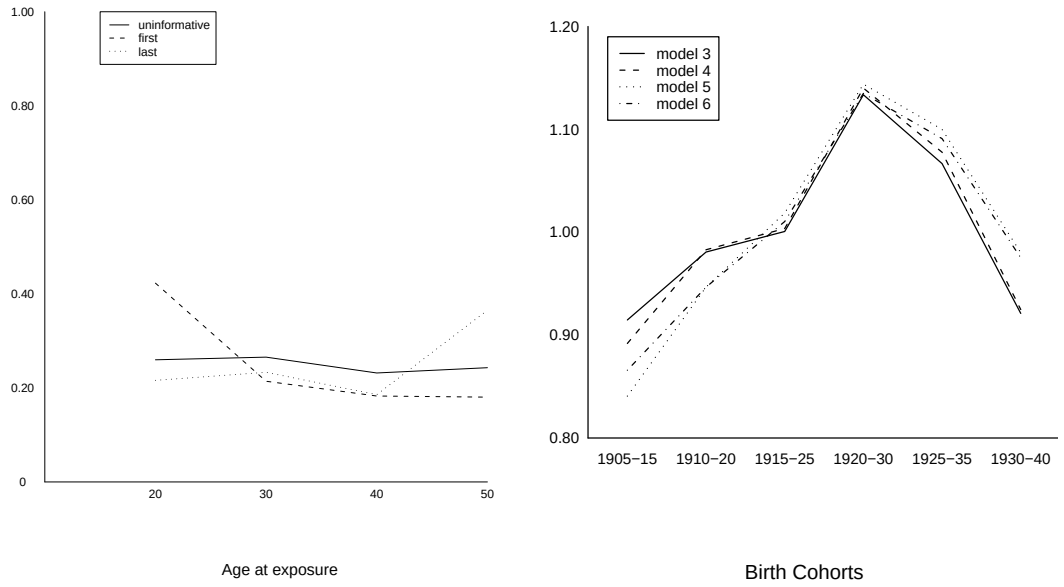
Table 9: β coefficients with their credibility interval (CI 90%) and age at exposure probabilities from model (6)

π_{20}	π_{30}	π_{40}	π_{50}
0.3986	0.2002	0.2001	0.2011

cohort	β_j
1905-15	-0.142 (-0.172,-0.112)
1910-20	-0.123 (-0.161,-0.087)
1915-25	-0.039 (-0.075,0.005)
1920-30	-0.050 (-0.089,-0.011)
1925-35	-0.031 (-0.066,0.004)
1930-40	0.049 (0.000,0.099)

Table 10: DIC measures for the space-time models

model	DIC	$\bar{D}$	$D(\bar{\theta})$	$\bar{D} - D(\bar{\theta})$
(1)	1958.389	1759.726	1561.064	198.6624
(3)	1958.226	1803.397	1648.498	154.8988
(4)	1945.535	1786.410	1627.284	159.1258
(5)	1887.251	1716.865	1546.479	170.3864
(6)	1969.187	1800.168	1631.149	169.0193



(a) (b)

Figure 4: (a) Age at exposure probability π_{jl} from model (5) when different prior are chosen. (b) Cohort effects from Bayesian models (3)-(6).

5. Conclusion and discussion

We have evaluated whether mortality differentials among areas, after removing space-time effects, could be associated with education. However, the relationship between putative individual cumulative exposure to risk agents and the spatial-temporal pattern of associated socioeconomic factors is complex: in the transformation process of an input map of socioeconomic conditions into an output map of cancer mortality, additional distortions cannot be excluded. Therefore much caution has to be used when interpreting ecological analysis such as that. As stated in section 3, the β coefficient quantifies the relationship between education score and mortality. In models (2), (5) and (6) we estimated a specific β_j for each birth cohort. An alternative formulation could be to consider β_j temporally (using a first-order random walk with independent gaussian increments) or spatially structured (for a review on varying coefficient models see Assunçao, 2003). Education index entered in this application without considering measurement errors on it. Models that take into consideration imprecisely observed covariates Bernardinelli *et al.* (1997) and uncertainty in imputed values are matter for future developments. In all the models described age-adjusted mortality rates are related to a predictor variable that is not age-adjusted. As Rosenbaum and Rubin (1984) pointed out this could produce biased estimates. The solutions adopted were a compromise to maintain a sufficient degree of simplicity in the model. More work is required on sensitivity analysis and model robustness (to the reader interested in problem, see Eberly and Carlin, 2000, Wakefield, 2003, Best *et al.*, 1999, Best *et al.*, 2005 and Knorr-Held and Rue, 2002). In particular allowing for residuals with spatial structure, instead of preventing confounding, could completely distort the direction of the association. In both model 5 and 6 the magnitude of the coefficients decrease over birth cohorts. This finding coincides with external evidence on a time trend toward greater homogeneity of risk for lung cancer among Tuscany's municipalities (see Lagazio *et al.*, 2003 and Vigotti *et al.*, 2001). The epidemics of lung cancer in males, in Tuscany, is now decreasing in all the municipalities with historically higher rates, while it is still increasing in the others, particularly in rural areas. The overall time trend is toward greater homogeneity among areas. An explanation of this pattern is that tobacco consumption was greater in the municipalities who underwent first industrialization and modernization. In the last decades of the twentieth century tobacco habits changed, with a general tendency to reduction in more developed areas. This migration of risk factors makes direct interpretation of the effect of socioeconomic factors more difficult. It is necessary to be aware of the dangers of over-interpretation of ecological analysis. Where after allowing for the biological effects of a confounding variable, there is a correlation between the residual geographical variability of the disease and the geographical distribution of the confounder, then the ecological analysis will wrongly estimate the confounder effect.

References

Assunçao R. (2003) Space varying coefficient models for small area data, *Environmetrics*, 14, 453–473.

Bernardinelli L., Pascutto C., Best N. and Gilks W. (1997) Disease mapping with errors in covariates, *Statistics in Medicine*, 16, 741–752.

Besag J., York J. and Mollié A. (1991) Bayesian image restoration, with applications in

spatial statistics, *Annals of the Institute of Statistical Mathematics*, 43, 1–59.
Best N., Arnold R., Thomas A., Waller L. and Conlon E. (1999) Bayesian models for spatially correlated disease and exposure data (with discussion), in: *Bayesian Statistics 6*, Oxford University Press: Oxford.
Best N., Richardson S. and Thomson A. (2005) A comparison of bayesian spatial models for disease mapping, *Statistical Methods in Mediacal Research*, 14, 35–59.
Cavelaars A., Kunst A. and et al. (2000) Educational differences in smoking: international comparison, *British Medical Journal*, 22, 1102–1107.
Clayton D. and Schifflers E. (1987) Models for temporal variation in cancer rates. i: age-period and age-cohort models, *Statistics in Medicine*, 6, 449–467.
Dreassi E. (2003) A space-time analysis of the relationship between material deprivation and mortality for lung cancer, *Environmetrics*, 14, 511–521.
Dreassi E., Biggeri A. and Catelan D. (2005) Space-time models with time dependent covariates for the analysis of the temporal lag between socio-economic factors and lung cancer mortality, *Statistics in Medicine, to appear, Published Online: 21 Feb 2005, DOI: 10.1002/sim.2063*.
Eberly L. and Carlin B. (2000) Identifiability and convergence issues for markov chain monte carlo fitting of spatial models, *Statistics in Medicine*, 19, 2279–2294.
Gamerman D. (1997) Sampling from the posterior distribution in generalized linear mixed models, *Statistics and Computing*, 7, 57–58.
Gelman A. and Rubin D. (1992) Inference from iterative simulation using multiple sequences (with discussion), *Statistical Science*, 7, 457–511.
Knorr-Held L. (2000) Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine*, 17-18, 2555–2568.
Knorr-Held R. and Rue H. (2002) On block updating in markov random fields for disease mapping, *Scandinavian Journal of Statistics*, 29, 597–614.
Lagazio C., Biggeri A. and Dreassi E. (2003) Age-period-cohort models on disease mapping, *Environmetrics*, 14, 475–490.
Lagazio C., Dreassi E. and Biggeri A. (2001) A hierarchical bayesian model for space-time variation of disease risk, *Statistical Modelling*, 1, 17–29.
Rosenbaum P. and Rubin D. (1984) Difficulties with regression analysis of age-adjusted rates, *Biometrics*, 40, 437–443.
Spiegelhalter D., Best N., Carlin B. and van der Linde A. (2002) Bayesian measures of model complexity and fit (with discussion), *Journal of the Royal Statistical Society B*, 64, 583–639.
Spiegelhalter D., Thomas A., Best N. and Gilks W. (2000) *WinBugs*, Medical Research Council Biostatistics Unit, Cambridge.
Townsend P., Phillimore P. and Beattie A. (1988) *Health and deprivation: inequalities and the north*, London: Croom Helm.
Vigotti M., Biggeri A., Dreassi E., Protti M. and Cislaghi C. (2001) *Atlas of mortality in Tuscany 1971-94*, Edizioni Plus: Università degli Studi di Pisa.
Wakefield J. (2003) Sensitivity analysis for ecological regression, *Biometrics*, 59, 9–17.